correspondence

TSG101 may be the prototype of a class of dominant negative ubiquitin regulators

The *TSG101* gene has been implicated in sporadic breast cancer[1]. Knockout of the murine orthologue resulted in fibroblast transformation and ability to form tumours[2]. The TSG101 protein contains a proline-rich region and a predicted coiled-coil domain, but no functional clues have been extracted from the sequence[1,2]. We show that the N-terminal domain of TSG101 belongs to a group of apparently inactive homologues of ubiquitin-conjugating enzymes (E2), which may be dominant negative regulators of E2 activity in cell-cycle control.

In database searches, non-globular protein domains produce numerous hits to proteins with biased composition, obscuring important but subtle signals[3,4]. When the non-redundant protein database at the NCBI was screened with the sequence of the TSG101 protein using the BLAST2 program[5], similarity was detected only to proline-rich and coiled-coil proteins. By contrast, the predicted N-terminal, globular[6] domain of TSG101 (134 amino acids) showed significant similarity ($P<10^{-4}$) to the hypothetical yeast protein YCL008c, a homologue of E2 ubiquitin-conjugating enzymes[7]. This search also detected similarity to several E2 sequences, but it was not statistically significant. YCL008c, however, is significantly similar to E2 ($P<10^{-3}$), establishing a link between TSG101 and ubiquitin-conjugating enzymes. This connection was supported by multiple alignment constructed using the MACAW program[8]. The alignment includes four blocks conserved in TSG101, YCL008c and E2, with $P<10^{-19}$ for three of them (Fig. 1). The active cysteine, which in E2 is conjugated with ubiquitin[9,10], is replaced by tyrosine in TSG101 and by alanine in YCL008c, though the surrounding motif is conserved (Fig. 1). Database searches with E2 sequences identified another yeast protein and a nematode protein similar to E2 but containing tyrosine or glutamine instead of the active cysteine (Fig. 1). The putative inactive E2 homologues contain an additional conserved motif, not shared by active ubiquitin-conjugating enzymes, which may be implicated in a common function (Fig. 1).

The three-dimensional structures of E2 enzymes from *Arabidopsis* and yeast are available[11,12]. The alignment of TSG101 with E2 includes all the structural elements of the latter, except for the two C-terminal α-helices, which are replaced by the proline-rich domain in TSG101 (Fig. 1). A statistical test[13] indicated high likelihood of structural similarity between TSG101 and the ubiquitin-conjugating enzymes ($P<10^{-7}$). A 3-D model for TSG101, based on the alignment with E2, was constructed and energy-minimized using the ICM method (Fig. 2)[14].

The ubiquitin system is implicated in several stages of cell-cycle control[15]. The ubiquitin-dependent degradation of the

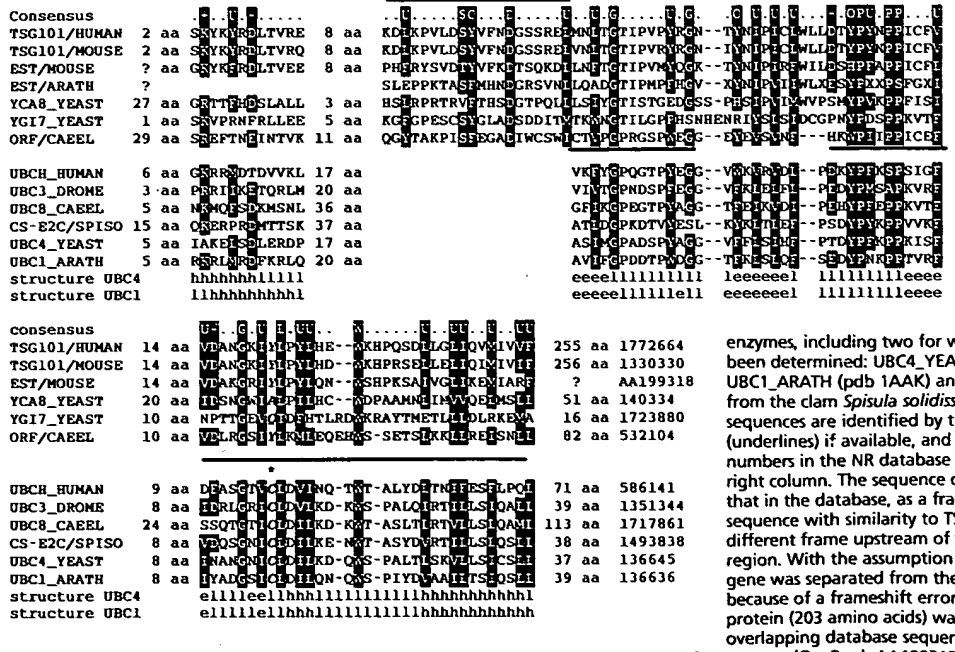

Fig. 1 Multiple alignment of TSG101 and other E2 derivatives with replaced active cysteine and bona-fide ubiquitin-conjugating enzymes. The upper block of seven sequences includes ubiquitin-conjugating enzyme homologues with replaced active cysteine, and the lower block of six sequences consists of diverse, active ubiquitin-conjugating enzymes, including two for which the 3-D structure has been determined: UBC4_YEAST (pdb 2UCE), UBC1_ARATH (pdb 1AAK) and a cyclin-specific enzyme from the clam *Spisula solidissima* (CS-E2C). The sequences are identified by their SWISS-PROT names (underlines) if available, and the gene identification numbers in the NR database are indicated in the far-right column. The sequence of YCL008c is different from that in the database, as a fragment of amino-acid sequence with similarity to TSG101 was detected in a different frame upstream of the annotated coding region. With the assumption that this portion of the gene was separated from the downstream portion because of a frameshift error, the putative complete protein (203 amino acids) was reconstructed from two overlapping database sequences (GenBank accessions S61879 and X59720). Two partial EST-encoded sequences with strong similarity to TSG101 from mouse (GenBank AA199318) and *Arabidopsis* (GenBank T76298) are included. The additional motif shared by TSG101 and other proteins with replaced active cysteine (but not the active E2 enzymes) is overlined. The distances between the aligned blocks and the distances from the protein termini are indicated by numbers. The consensus includes amino acid residues conserved in most of the aligned sequences; the residues that conform with the consensus are shown by reverse typing. U, bulky hydrophobic residue (I,L,V,M,F,Y,W); O, aromatic residue; S, serine or threonine; –, negatively charged residue; +, positively charged residue; dots indicate any residue. The ubiquitin-conjugating cysteine is indicated by an asterisk. The alignment downstream of the active cysteine was modified manually to align the tryptophan residue, which is conserved in all available sequences of E2 enzymes. The experimentally determined secondary structure elements for UBC4_YEAST and UBC1_ARATH are shown below the alignment; h, α-helix; e, extended conformation (β-sheet); l, loop.

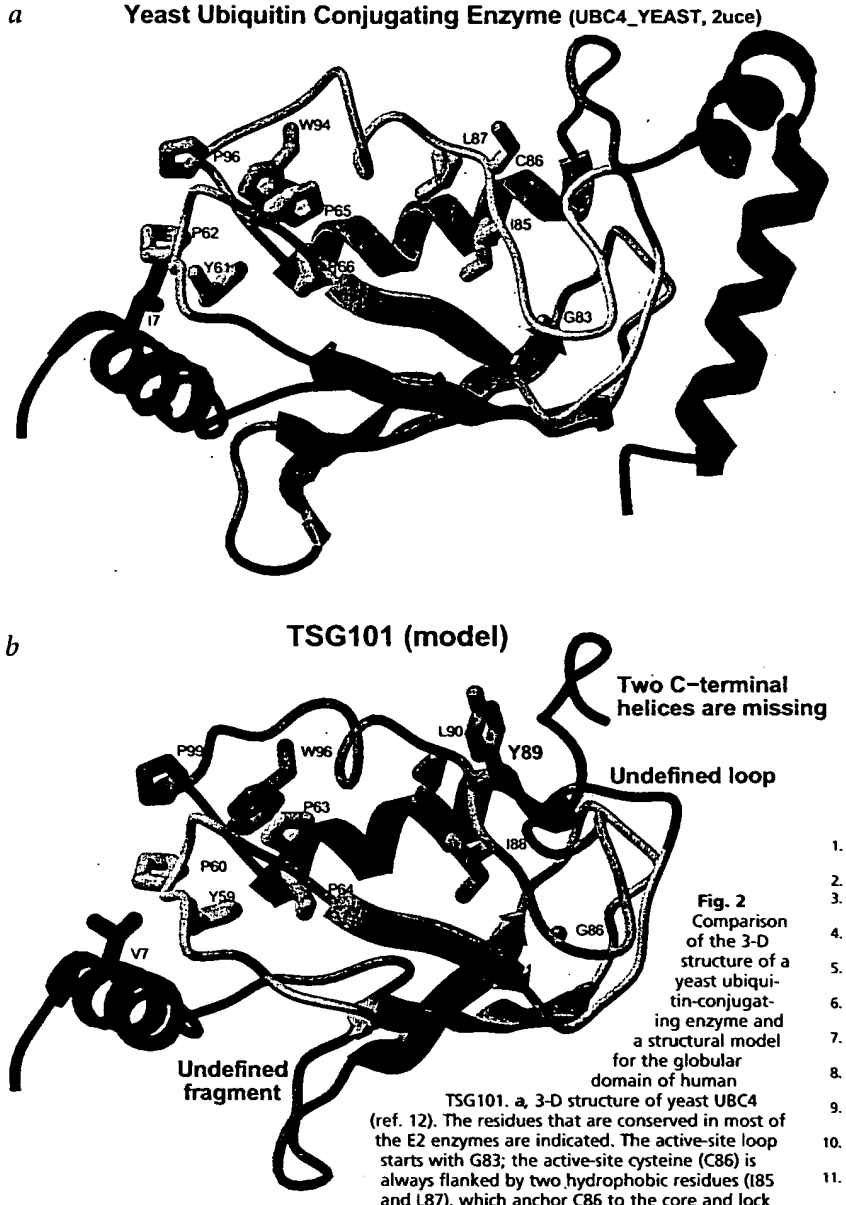

Fig. 2 Comparison of the 3-D structure of a yeast ubiquitin-conjugating enzyme and a structural model for the globular domain of human TSG101. *a*, 3-D structure of yeast UBC4 (ref. 12). The residues that are conserved in most of the E2 enzymes are indicated. The active-site loop starts with G83; the active-site cysteine (C86) is always flanked by two hydrophobic residues (I85 and L87), which anchor C86 to the core and lock the orientation of its side chain. The loop between C86 and W94, a residue conserved in all known E2 sequences, may vary in length in different groups of enzymes (Fig. 1). *b*, Hypothetical model of the TSG101 structure based on the alignment in Fig. 1 and the UBC4 structure. The model was free of steric clashes (van der Waals energy –608 kcal/mol) and contained only one medium size (63 Å^3) cavity. All the essential residues spatially adjacent to the active cysteine in E2 are conserved in TSG101. The conserved tryptophan and the adjacent residues form a hydrophobic cluster supporting the active site geometry. Thus, the active site loop in TSG101 is likely to have the same conformation as in E2 despite the replacement of the active cysteine. The two regions that could not be reliably aligned (Fig. 1) and the two C-terminal helices, which are apparently missing in TSG101, do not contact the active site and are unlikely to affect ligand binding. Despite the low sequence similarity in the N-terminal helix, a conserved hydrophobic contact (I7-P62 and V7-P60 in UBC4 and TSG101, respectively) could be proposed. A colour gradient from blue (N terminus) to red (C terminus) is used in both images. The images were generated with ICM software[25].

cyclin-dependent kinase inhibitor SIC2 is critical for the G1-S transition in yeast[16]. Ubiquitin-mediated destruction of the anaphase inhibitor PDS1 and of cyclin B is required for the metaphase-anaphase progression and for the telophase-G1 progression, respectively[17–19]. The checkpoint proteins p53 and NF-κB are also degraded by the ubiquitin cascade[20,21]. Cyclin-specific E2 enzymes have been isolated[22,23]; their derivatives with serine substituted for the active cysteine are dominant negative inhibitors of cyclin destruction and block cells in metaphase[24]. TSG101-like proteins may belong to a novel class of cell-cycle regulators, intercepting the ubiquitin cascade by forming non-productive complexes with ubiquitin, target proteins or other components of the ubiquitin machinery.

Note added in proof: We have recently become aware of similar conclusions independently reached by others (Ponting, C.P., Cai, Y.-D., Bork, P.J. *Mol. Med.*, in the press).

Eugene V. Koonin[1] & Ruben A. Abagyan[2]

[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda Maryland 20894, USA. [2]Courant Institute of Mathematics and Skirball Institute of Biomolecular Medicine, New York University, New York, New York 10016, USA. Correspondence should be addressed to E.V.K. e-mail koonin@ncbi.nlm.nih.gov

1. Li, L., Li, X., Francke, U. & Cohen, S.N. *Cell* **88**, 143–154 (1997).
2. Li, L. & Cohen, S.N. *Cell* **85**, 319–329 (1996).
3. Altschul, S.F., Boguski, M.S., Gish, W. & Wootton, J.C. *Nature Genet.* **6**, 119–129 (1994).
4. Koonin, E.V., Altschul, S.F. & Bork, P. *Nature Genet.* **13**, 266–268 (1996).
5. Altschul, S.F. & Gish, W. *Meth. Enzymol.* **266**, 460–480 (1996).
6. Wootton, J.C. & Federhen, S. *Meth. Enzymol.* **266**, 554–571 (1996).
7. Koonin, E.V., Bork, P. & Sander, C. *EMBO J.* **13**, 493–503 (1994).
8. Schuler, G.D., Altschul, S.F. & Lipman, D.J. *Proteins Struct. Funct. Genet.* **9**, 180–190 (1991).
9. Sung, P., Prakash, S. & Prakash, L. *J. Mol. Biol.* **221**, 745–749 (1991).
10. Sullivan, M.L. & Vierstra, R.D. *J. Biol. Chem.* **266**, 23878–23885 (1991).
11. Cook, W.J., Jeffrey, L.C., Sullivan, M.L. & Vierstra, R.D. *J. Biol. Chem.* **267**, 15116–15121 (1992).
12. Cook, W.J., Jeffrey, L.C., Xu, Y. & Chau, V. *Biochemistry* **32**, 13809–13817 (1993).
13. Abagyan, R.A. & Batalov, S.N. *J. Mol. Biol.* (in press).
14. Abagyan, R.A., Totrov, M.M. & Kuznetsov, D.A. *J. Comp. Chem.* **15**, 488–506 (1994).
15. King, R.W., Deshaies, R.J., Peters, J.M. & Kirschner, M.W. *Science* **274**, 1652–1659 (1996).
16. Schwob, E., Bohm, T., Mendenhall, M.D. & Nasmyth, K. *Cell* **79**, 233–244 (1994).
17. Cohen-Fix, O., Peters, J.M., Kirschner, M.W. & Koshland, D. *Genes Dev.* **10**, 3081–3093 (1996).
18. Glotzer, M., Murray, A.W. & Kirschner, M.W. *Nature* **349**, 132–138 (1991).
19. Holloway, S.L., Glotzer, M., King, R.W. & Murray, A.W. *Cell* **73**, 1393–1402 (1993).
20. Scheffner, M., Huibregtse, J.M., Vierstra, R.D. & Howley, P.M. *Cell* **75**, 495–505 (1993).
21. Palombella, V.J., Rando, O.J., Goldberg, A.L. & Maniatis, T. *Cell* **78**, 773–785 (1994).
22. Yu, H., King, R.W., Peters, J.M. & Kirschner, M.W. *Curr. Biol.* **6**, 455–466 (1996).
23. Aristarkhov, A. *et al. Proc. Natl. Acad. Sci. USA* **93**, 4294–4299 (1996).
24. Townsley, F.M.A. *et al. Proc. Natl. Acad. Sci. USA* **94**, 2362–2367 (1997).
25. Molsoft, L.L.C. *ICN Software Manual*, version 2.6 (http://www.molsoft.com) (1996).